

Corpus Annotation Guidelines

This document contains the Coreference Corpus Annotation Guidelines for ACL Anthology papers.

The last section describes [How to Annotate with MMAX2](#).

The corpus is described in the following paper:

Ulrich Schäfer, Christian Spurk, Jörg Steffen: *A fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology*. Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012), pages 1059-1070, Mumbai, India, 2012. PDF URL: <http://www.aclweb.org/anthology/C12-2103>.

```
@InProceedings{schafer-spurk-steffen:2012:POSTERS,
  author   = {Schäfer, Ulrich and Spurk, Christian and Steffen, Jörg},
  title    = {A Fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology},
  booktitle = {Proceedings of COLING 2012: Posters},
  month    = {December},
  year     = {2012},
  address  = {Mumbai, India},
  publisher = {The COLING 2012 Organizing Committee},
  pages    = {1059--1070},
  url      = {http://www.aclweb.org/anthology/C12-2103}
}
```

The annotated data is available at take.dfki.de.

1. [Terminology](#)
 1. [Coreference](#)
 2. [Markable](#)
 3. [Coreference Set](#)
 4. [Antecedent](#)
 5. [Related Terminology in the Literature](#)
2. [Markables to Annotate](#)
 1. [Named Entities](#)
 2. [Definite Noun Phrases](#)
 3. [Indefinite Noun Phrases](#)
 4. [Conjunctions](#)
 5. [Personal Pronouns](#)
 6. [Possessive Pronouns](#)
 7. [Reflexive Pronouns](#)
 8. [Demonstrative Pronouns](#)
 9. [Relative Pronouns](#)
3. [Markables Not to Annotate](#)
 1. [Only Noun Phrase and Possessive Determiner Markables](#)
 2. [No Relative Clauses](#)
 3. [Only Definite Predicate Nominatives with Definite Subjects](#)
 4. [Only Direct Anaphora](#)
 5. [No Bound Anaphora](#)
4. [General Annotation Principles](#)
 1. [Strive to Find Maximal Coreference Sets](#)
 2. [Annotate Maximally Long Markables](#)
 3. [Only Contiguous Markables](#)
 4. [Look Out for Nested Markables](#)
 5. [Review and Reuse Pre-Annotations](#)
 6. [Time, Space and Modality in References](#)
 7. [Tips on Deciding between NP and NE Markable Types](#)
 8. [Miscellaneous Tips and Clarifications](#)
5. [How to Annotate with MMAX2](#)
 1. [Installation](#)
 2. [Opening MMAX2 Projects](#)
 3. [MMAX2 Windows](#)
 1. [The Markable Level Control Panel](#)
 2. [The Main Window](#)
 3. [The Attribute Window](#)
 4. [Basic Markable Operations](#)
 1. [Creating a Markable](#)

2. [Deleting a Markable](#)
3. [Resizing a Markable](#)
5. [Modifying Markable Relations](#)
 1. [Adding Markables to Coreference Sets](#)
 2. [Removing Markables from Coreference Sets](#)
6. [Saving the Annotation](#)

Terminology

This section describes general terminology used in this annotation guide. Some of the terms are used with small differences in the literature. Thus, this section is also meant to be a disambiguation section for the relevant terminology. If you are interested in how the terminology used in this guide differs from the terminology used in the literature, you may read the optional section [“Related Terminology in the Literature”](#).

Coreference

We define **coreference** to be the act of two or more textual representations (words, phrases, etc.) to refer to the same entity – or more generally to the same referent – in the real world. For example, in the sentence *Peter likes dogs; he has three of them.* the noun phrase (NP) *Peter* and the pronoun *he* refer to the same real world entity which is the person who is called Peter; thus, *Peter* and *he* are coreferential. Similarly, *dogs* and *them* are coreferential as both phrases refer to the dog concept. Another example would be the sentence *Today it is the 10th of July 2008.* where *today* and *the 10th of July* are coreferential as they (presumably) both refer to the same date.

In general, coreference is not restricted to textual representations. However, as we are working with texts only in this annotation task, we restrict our notion of coreference to textual representations.

Coreference is contextually dependent: two phrases may be coreferential in one sentence but they might not be coreferential in another sentence. For example, in *Peter is my father.* the NP *Peter* is coreferential with the NP *my father*, however, in *I introduced Peter to my father.* the two NPs are not coreferential.

For our corpus annotation purposes, we are only interested in certain coreferential items; the upcoming section on the [markables to annotate](#) describes these items in more detail.

Markable

The representations (words, phrases, etc.) in a text which can potentially be coreferential are called **markables**. For example, in the sentence *Mary and Peter are going to the cinema.* we have quite a few markables:

1. *Mary*
2. *Peter*
3. *Mary and Peter*
4. *the cinema*
5. *going to the cinema*
6. *Mary and Peter are going to the cinema*

Perhaps this might seem to be too many. When considering different follow-up sentences it should become clear, though, that these markables are indeed all possible (coreferential markables underlined):

1. *She* likes that.
2. *He* likes that.
3. *They* often do this.
4. *It's* the only cinema in town.
5. *They* often do *this*.
6. Luckily *Mary's dad* is unaware of *that*.

For our corpus annotation purposes, however, we are only interested in a small portion of all possible markables; the upcoming section on the [markables to annotate](#) describes these markables in more detail.

In the literature, markables are sometimes also called *mentions*.

Coreference Set

Under a **coreference set** we understand sets of markables of a document that are all coreferential.

As an example let's consider this short document:

- *Peter and Mary are going to the cinema. They often do that. Sometimes Peter's colleague, John Baker, joins them.*

Here are all (relevant) coreference sets that can be found in the example document:

- { *Peter, Peter* }
- { *Peter and Mary, They, them* }
- { *going to the cinema, that* }
- { *Peter's colleague, John Baker* }

Coreference sets with just one markable are uninteresting and can therefore usually be ignored (such as in the previous example the set { *Mary* }). In the literature, coreference sets are sometimes also called *mention chains*.

Antecedent

A markable with which another markable is coreferential is called the **antecedent** of the other markable.

Example: in the sentence *Peter likes dogs; he has three of them.* the markable *he* has the antecedent *Peter*.

Related Terminology in the Literature

NOTE: This section is purely informational and should not need to read it for the annotations. It is only there to relate our terminology to frequently used terminology from the literature. Thus, the terminology presented in this section is not relevant for the annotations!

In the literature, coreference is often talked about in conjunction with **anaphora**. Compared to our notion of coreference, anaphora is a very similar linguistic phenomenon which is also about textual reference, however, without requiring the involvement of real-world entities. Anaphora only deals with intra-textual reference, sometimes also called (purely) syntactic reference. For example, the word *today* may be coreferential with the current date, however, without taking this extra-textual entity – the date – into account, it is not possible to find intra-textual coreference. Thus, the set of all anaphoric phenomena is a superset of all coreference phenomena.

An **anaphor** is then a markable that refers back to some other markable with which it stands in an anaphoric relation. A **cataphor** is a markable which makes a forward reference to some other markable with which it stands in an anaphoric relation.

Examples:

- *Peter likes dogs; he has three of them.*
→ The markable *he* is an anaphor as it refers to (and even is coreferential with) the previous markable *Peter*.
- *Because he had already been feeling sick all day, Peter eventually went to the doctor.*
→ The markable *he* is a cataphor as it refers to (and even is coreferential with) the following markable *Peter*.
- *Peter had long planned to go; which he eventually did yesterday.*
→ Anaphora example which bears no coreference: the verb *did* is an anaphor referring to the verb *go*; however, neither of the two verbs refers to a real world referent, i.e., there is no coreference involved here.

In the literature, **coreference** is actually not always restricted to textual reference. This restriction is mainly a simplification for easing our annotations. Actually, coreference and anaphora can be viewed as two sets whose intersection defines our notion of coreference.

Markables to Annotate

In general for our annotation task, we are only interested in coreferential **noun phrase (NP) and possessive determiner markables** (as defined [above](#)). This especially means that you can ignore all other markables, i.e., all such markables which are none of the types specified in the following sections. Don't worry, though, about superfluously annotated markables: we will automatically remove all markables later which do not belong to any coreference sets. So you don't have to remove markables again that you might

accidentally create during the annotation.

In this section we describe the different kinds of markables that shall be annotated in our annotation task. Note, however, that [the general annotation principles](#) we will describe in the section following this one always take precedence over the upcoming descriptions.

Named Entities

Names and named entities (NEs) are usually (definite) NPs and as such can enter into coreference relations, i.e., they are relevant markables. NEs may be – among others – names of companies, organizations, persons, locations, languages, currencies, bands, programming languages, standards, scientific fields, language names, etc. As a special case for our ACL Anthology corpus, we consider citations in scientific papers as NEs, too.

Examples (NE markables underlined):

- *Peter lives in Manhattan.*
- *Mr. Anderson is a good programmer in Perl.*
- *This approach has been criticized by van Deemter and Kibble (2000).*
- *Especially extracting information from abstracts of MEDLINE is regarded as one of the most important tasks (Strazli et al., 2001; Tendon and Byrne, 2000; Yono, 2001).*
- *(Tsuruji et al., 2004) have shown that the improvements with TnT are negligible.*
- *The German Wikipedia is the second largest Wikipedia today.*
- *Information Extraction (IE) and Machine Translation (MT) are popular areas of Computational Linguistics.*
- *German is a nice language.*

Variable names are – as the name implies – NEs, too. Note, however, that variable names may frequently change their meaning in different contexts. So, deciding whether two variables are coreferential or not requires close examination of the context. Concerning the annotation of variable names, see also [the tip below](#).

In the context of NEs nested within other NEs or within NPs you are advised to read [the note below](#). As it may be difficult to differentiate between NE markables and normal NP markables, we have compiled some tips that help you in [deciding between NE markables and other NP markables](#).

See also the corresponding [Wikipedia article](#).

Definite Noun Phrases

Definite noun phrases are NPs which correspond to a specific and identifiable entity in a given context. In many cases this definiteness is marked by the definite article “the” or a demonstrative determiner such as “these” or “that”.

Examples (definite NP markables underlined):

- *The man likes his cat.*
- *These men like cats.*
- *[...] In such a context they felt that the game would not be fun anymore.*
- *On the table there was a flower and a vase. It looked as though someone had prepared the vase for the flower.*
- *[...] Both approaches have their own advantages.*
- *Jane and Andy love each other.*
- *My mother knows someone who is really animal-loving. This man is especially fond of caterpillars.*

Pronouns and named entities are definite NPs, too, however we mark them separately for our annotation task.

Indefinite Noun Phrases

Indefinite noun phrases are NPs which do not correspond to a specific and identifiable entity in a given context. In many cases this indefiniteness is marked by the indefinite articles “a” and “an” or it is indicated by the lack of a certain determiner.

Examples (indefinite NP markables underlined):

- *Every man loves a woman.*

- *These men like cats.*
- *Cats love fish.*
- *On the table there was a flower and a vase. It looked as though someone had prepared the vase for the flower.*

Conjunctions

For our annotation task, we define a conjunction to be an NP which results by conjoining other NPs. The most common junctor which is used for conjoining NPs is “and”. Other junctors include, for example, “or”, “as well as” or the discontinuous junctor “both ... and”.

Examples (conjunction markables underlined):

- *Jane and Andy love each other.*
- *Peter, Paul, Mary and Jane go to the cinema.*
- *Eggs and flour are important ingredients for apple pies and pancakes.*
- *My brother and I invited Mary for dinner yesterday. Although my brother usually doesn't like the fish we prepared, he and Mary enjoyed the meal as well as the whole evening very much.*
- *Both Peter and Mary have already been in Vietnam.*

Examples of coreferential conjunctions (conjunction and coreferential markable underlined):

- *Peter and Mary are happy with their life.*
- *Although they didn't know why, Peter and Mary somehow felt sick.*
- *Peter, Mary and John make up the Smith family.*
- *Both HPSG and LFG are popular grammar formalisms with their own supporters.*

Personal Pronouns

Personal pronouns are pronouns which stand for another NP and which are even complete NPs themselves. The most common personal pronouns in English are “I”, “you”, “he”, “she”, “they”, “it”, “me”, “him”, “us”, “them”, “her” and “we”.

Examples (personal pronoun markables underlined):

- *Mary and I like watching horror films although she is often afraid of them.*
- *The accident was terrible. But although he kept thinking it was his fault Peter didn't give up. (NB: it is assumed to be coreferential with *The accident* here)*

Watch out for expletive pronouns, though: they are not referring to anything and thus they are no markables. Examples (expletive pronouns underlined):

- *It's hard to say whether it is important to see this movie.*
- *It seems as if it was raining.*

See also the corresponding Wikipedia articles on [personal](#) and [expletive](#) pronouns.

Possessive Pronouns

Possessive pronouns in a strict sense are NPs which stand for another NP and which attribute ownership to the NP they substitute, e.g., “mine”, “hers” or “ours”. In our annotation task, we assume a broader sense in which possessive determiners are also considered to be possessive pronouns, e.g., “his”, “her” or “my”.

Examples (possessive pronoun markables underlined):

- *I have borrowed Mary my book because she had lost hers ages ago.*
- *Peter visits his mom and dad.*
- *Our approach shows that theirs cannot be right.*
- *Her father enjoys making his own films.*

See also the corresponding [Wikipedia article](#).

Reflexive Pronouns

Reflexive pronouns are pronouns that substitute the NP to which they refer in the same clause as the NP. The most common reflexive pronouns in English are “myself”, “yourself”, “himself”, “herself”, “themselves”,

“itself”, “ourselves”, “yourselves” and “themselves”.

Examples (reflexive pronoun markables underlined):

- *Peter washes himself.*
- *The tourists themselves remained unhurt.*

See also the corresponding [Wikipedia article](#).

Demonstrative Pronouns

In our annotation task, demonstrative pronouns are pronouns which are NPs that stand for some other NP of the discourse. As such they are very similar to [personal pronouns](#). The most common demonstrative pronouns in English are “this”, “that”, “these” and “those” while for the annotations in our corpus you will mostly find only “these” as a markable.

Examples (demonstrative pronoun markables underlined):

- *Before going to bed I always read sci-fi books. I simply love these.*
- *It will show the date of the event, and, if these can be found, the location and the organizer of it.*

See also the corresponding [Wikipedia article](#).

Relative Pronouns

Relative pronouns are pronouns which introduce relative clauses. We are only interested in relative pronouns that introduce non-restrictive relative clauses (see also [Wikipedia article](#)). In restrictive relative clauses, the relative pronoun does not refer to any real-world entity and therefore it can't be a coreferential markable. (The relative pronoun refers syntactically to the head noun of the noun phrase (NP) to which the relative clause belongs, however, this is no coreference; the NP is semantically incomplete without the relative clause.) In non-restrictive clauses, the relative pronoun really corefers with the noun phrase (NP) to which the relative clause belongs.

The relative pronouns in English which are also relevant for your annotations are “who”, “which”, “whose”, “where”, “whom” and “when” as well as sometimes “that” and rarely “why”.

Examples (relative pronoun markables underlined):

- *John saw Mary, who he immediately hugged.*
- *Mary loves the Eiffel Tower, which she already visited more than 20 times.*
- *Mary lives in Berlin, where she knows a lot of nice people.*
- *Mary also knows Jimmy, whom she regularly sees at work.*
- *Mary will come next Thursday, when John is still away on business.*

Summing up, before you annotate a relative pronoun as a markable, you must make sure that it introduces a non-restrictive relative clause. If it should introduce a restrictive relative clause, then you must not annotate it.

Side note: because the relative clause that the pronoun introduces in a restrictive relative clause belongs to the preceding NP, you must annotate the relative clause as part of the NP, in order to obey [the rule of annotating maximally long markables](#). In non-restrictive relative clauses, the relative clause does not belong to the preceding NP and therefore must not be annotated as a part of this NP.

See also the corresponding [Wikipedia article on relative pronouns](#).

Markables Not to Annotate

This section is just meant to further clarify and elaborate on the annotation objective given above: only annotate noun phrase (NP) and possessive determiner markables which are coreferential with some other markable. The concrete markables to annotate have already been described in the [“Markables to Annotate”](#) section. This section is meant to aid in deciding whether a markable is really relevant or not.

Only Noun Phrase and Possessive Determiner Markables

As may be guessed from the examples so far, the majority of markables are noun phrases (NPs). If you find an NP in some document, then it is well possible that it is also a markable. The same holds for possessive determiners like *my* and *her* in *I like my shoes but not her skirt.*: possessive determiners are always markables, too. Other kinds of markables are not so frequent and are sometimes even difficult to track down. Because of this we have decided to only annotate NP markables and possessive determiner markables and ignore all kinds of other markables.

Here are some examples of markables which should thus be ignored (some of the markables in question are underlined):

- *Today we have finally won the prize!*
- *Peter and Mary are going to the cinema. They often do that.*
- *The dogs are playing in the garden. There they also have all their toys.*
- *Mary and Peter are going to the cinema. Luckily Mary's dad is unaware of that.*
- *The Spanish text is longer than the German.*
- *We use Information Extraction techniques for this task.*

Together with *mine*, *yours*, *hers*, etc. possessive determiners are often (sloppily) called possessive pronouns in the literature – and in this guide.

No Relative Clauses

For our annotation task, we are not interested in relative clause markables. So you should *not* have any markables in your annotations like the ones underlined in the following examples:

- *Peter is the manager who enjoys playing football.*
- *This is the book that I bought last week.*
- *These are the shoes which I like best.*
- *These are the shoes I like best.*

NB: it is still perfectly valid to have relative clauses annotated as part of some other markables like in the following examples (coreferential markables underlined):

- *Peter is the manager who enjoys playing football.*
- *Here is the book that John bought last week for Jane's mother in law. I have already read it, too.*
- *These are the shoes I like best.*

Only Definite Predicate Nominatives with Definite Subjects

A predicate nominative can only be a coreferential markable if it is definite and connected to a definite subject. Thus, in the following examples only the underlined markables are relevant:

- *A mason is a workman.* (both NPs are not definite)
- *Peter is a mason.* (the second NP is not definite)
- *Peter is Mary's brother.* (both NPs are definite)
- *Peter is one of Mary's brothers.* (the second NP is not definite)
- *Our mason is a nice guy.* (the second NP is not definite)
- *Our mason is the best.* (both NPs are definite)

Only Direct Anaphora

Let's have a look at some phenomena which are similar to coreferences but which should not be annotated although these phenomena are even sometimes called coreferential in the literature. Consider the following example sentence which contains different kinds of coreference or at least similar phenomena:

- *Peter has painted his room red. He has left the door yellow, though.*

These are the coreference sets one might come up with for this example:

- { *Peter*, *He* }
- { *his room*, *the door* }

It is obvious that in the first set *He* is a coreferential markable with the antecedent *Peter* as both markables refer to the same entity. In the second set, *the door* might be considered to be a markable with the antecedent *his room*; both markables in some way refer to the same entity – a room usually has a door –, however, they denote different things. The reference we see here is called *indirect anaphora* or

bridging reference in the literature. According to our strict definition [above](#), the two markables are not coreferential as they do not directly refer to the exact same entity, though. Thus, you should not annotate such phenomena in our corpus. We are only interested in what is called *direct anaphora* in the literature (cf. for example in Mitkov 2003, p. 269).

Here are further examples for bridging references and similar references (antecedents and indirectly anaphoric markables underlined):

- *The bar is crowded. The waitress is stressed out.*
- *Mary likes Peter's dress style. She especially likes his red trousers.*
- *My fan doesn't work anymore. The power cable is torn out.*
- *Every student wants to have good study conditions. But they don't want any tuition fees.*
- *Many people love travelling, especially those who have already travelled with their parents. (*those* is not coreferential with the *Many people* group, it only instantiates this group)*
- *We need to prepare for the expected impacts as such an impact might have grave consequences. (*such impacts* is not coreferential with *the expected impacts* group, it only instantiates this group)*

If you find similar examples in our corpus, then just ignore them as they bear no coreference as we understand it in our annotation task.

No Bound Anaphora

Bound anaphora (Mitkov 2003, p. 268) is a reference phenomenon which is similar to our notion of coreference. However, in bound anaphora, referents are not coreferential and thus they must not be annotated in our annotation task. Examples of bound anaphora (non-coreferential referents underlined):

- *Every teacher likes his job.*
- *No dog will ever drive a car. They just can't do that.*
- *Some birds can't fly although they have wings.*

The reason why referents are not coreferential in these examples lies in the first referent of each example: these referents *do not refer to a specific entity* in the real world.

It is often difficult to tell coreference and bound anaphora apart. The core question you always need to answer in order to decide between coreference and bound anaphora is to judge whether a referent is referring to a specific entity in the real world or not. Examples (with difficult referents underlined):

- *Most people brought their drinks themselves.*
→ Does *Most people* perhaps refer to a specific entity, e.g., a specific group of people?
- *Some dogs looked so sad that I bought them.*
→ *Some dogs* probably refers to those dogs that were bought, i.e., it is a markable.
- *Some dogs look so sad that people hug them.*
→ *Some dogs* appears to not be a specific group of dogs to which the speaker would refer, i.e., it is not a markable.
- *There are 523 people in the park.*
→ *523 people* refers to a fixed group of people, i.e., it is a markable.
- *There are about 500 people in the park.*
→ *about 500 people* does not refer to a fixed group of people, i.e., it is not a markable.
- *One could say that HPSG is better than LFG.*
→ *One* probably does not refer to a specific person entity, i.e., it is not a markable.

If you are really in doubt whether an example in the annotations is coreferential or not, then rather don't annotate it.

General Annotation Principles

Annotation of corpora is a broad field; there are many things that could be annotated. As you may have guessed already from the terminology section, though, the whole annotation of our corpus is only about annotating coreferences. This section is about general principles that shall guide you throughout your annotation task. No matter what particular kind of [coreference phenomenon](#) you are annotating, you should always keep these principles in mind when deciding whether a particular coreference shall be annotated or not and how this coreference shall be annotated.

Strive to Find Maximal Coreference Sets

Whenever you find two coreferential markables, then only create a new coreference set if there is not any other coreference set, yet, whose elements refer to the same referent as the two newly found markables. For every pair of document and real-world referent there should be only one coreference set with markables of the document referring to the referent. You should especially pay attention to this when finding new coreferential markables towards the end of the document.

Take this document as an example:

- *Recently I have bought a new computer which had Microsoft Windows preinstalled. As I wanted to have Windows anyway, I was happy for not having to install it myself. But then the problems began. I could neither [skipping pages of problem descriptions] After all these problems with Windows I was happy to discover GNU/Linux. It was a good feeling to eventually trash the Microsoft OS.*

In the first part of the document – among others – you should find this coreference set:

- { *Microsoft Windows, Windows, it* }

When you reach the end of the document, after having annotated lots of other markables, you encounter the two coreferential markables *Windows* and *the Microsoft OS*. Instead of creating a new coreference set for them you should add them to the existing set for Windows:

- { *Microsoft Windows, Windows, it, [maybe other coreferential markables from the skipped pages], Windows, the Microsoft OS* }

■ Annotate Maximally Long Markables

Every markable you annotate should be maximally long. For example, when you annotate noun phrase (NP) markables, you should make sure to start your annotation with any determiner of the NP and include any adjectives modifying the NP. Also, if there is any phrase modifying the NP you should also include this phrase in the markable. Such modifying phrases are usually prepositional phrases (PPs) or restrictive relative clauses. Parenthetical phrases – or parenthetical modifiers in general – are other common post-modifiers in NPs.

Here are some examples of maximally long markables:

- *Mary saw the man who stole the manager's car. Nonetheless, he was not caught.*
- *Mary saw the boy playing in the garden from her window. He looked funny.*
- *Mary saw the boy playing in the garden with his dog from her window. He looked funny.*
- *Mary saw the boy playing in the garden with the curtain from her window. She threw an angry look at him.*
- *Mary carefully took the bucket with water from the hose she had connected to the wall before. It was full to the brim.*
- *Mary carefully took the bucket with water flowing over – she had filled it up too much.*
- *Mary took the little yellow plastic duck. It squeaked.*
- *Mary takes the little yellow plastic duck she had bought for her daughter yesterday. It squeaks.*
- *Mary took the little yellow plastic duck she had bought for her daughter yesterday. It squeaked.*
- *Sen. John Smith (R., Arizona) visited the company last week. He spoke at length there.*
- *Mary bought the little yellow plastic duck (called Jimmy). The duck squeaked.*
- *Mary bought the little yellow plastic duck (which she called Jimmy). The duck squeaked.*
- *Mary bought the little yellow plastic duck (and called it Jimmy). The duck squeaked.*
- *Mary visits Jim, who has just returned from holidays.*
- *John lives near the city of Paris, where he works at PSA.*

As can be seen from the examples, you always have to consider the context of the current sentences in order to determine whether certain words belong to a markable or not. This is especially true for adverbials which might modify both some part of the markable or the verb of the current sentence. Non-restrictive relative clauses are not part of the preceding NP; restrictive relative clauses are. Relative clauses without a relative pronoun are almost always restrictive relative clauses and must therefore be annotated as part of the markable.

Special note on citation post-modifiers which can often be found in our corpus: citations in parentheses are parenthetical modifiers, too. As such, if they relate to the preceding noun, they must be part of a markable for the corresponding NP. Example: *We used the PaSY parser (Smith et al., 2007).*

■ Only Contiguous Markables

A markable can be noncontinuous, i.e., between its parts can be other words which do not belong to the

markable. You can safely ignore such markables in your annotations. As noun phrases cannot be noncontinuous in English this rule can be seen as a specialization of [the guideline to only annotate NPs and possessive determiners](#).

Examples (parts of the noncontinuous markables underlined):

- *Peter met Mary at the university. They talked for hours.*
- *Mary brought her kids to the kindergarten. They came by car.*

In the literature, such cases are sometimes called *split-antecedent anaphora*.

Look Out for Nested Markables

Markables can be nested in other markables, so make sure to look out for such cases. In theory, markables can be nested arbitrarily deep, however, you will probably not find markables that are nested deeper than three levels.

Consider this example document:

- *ACME Inc. has many employees working with Python. Peter is the company's guru for Python, although he only started to use the programming language when he joined ACME.*

These are all the coreference sets for the document that we would be interested in for our annotation task:

- { *ACME Inc., the company, ACME* }
- { *Python, Python, the programming language* }
- { *Peter, the company's guru for Python, he, he* }

The interesting bit about this example is that the markable *the company's guru for Python* contains two other markables *Python* and *the company*.

Here is another example with markables nested up to three levels deep:

- *The ACME tree-bank format is the format of the tree-bank which has been developed by ACME Inc. The tree-bank of the company contains 12,000 sentences.*

These are all coreference sets for the example document that we would be interested in for our annotation task:

- { *The ACME tree-bank format, the format of the tree-bank which has been developed by ACME Inc.* }
- { *the tree-bank which has been developed by ACME Inc., The tree-bank of the company* }
- { *ACME Inc., the company* }

In this example you should take particular note of the nesting of named entities (NEs): the NE *ACME* of the first sentence is not part of any coreference set, although it is a coreferential NE. However, in this context the NE is **not a noun phrase (NP)** but simply a modifier noun. For our annotation task, we are only interested in NPs, so nested NEs should only be annotated as markables when they are used as NPs. If you are unsure, whether some NE is used as an NP or not, then try to replace the NE with a full NP to see whether the sentence remains grammatical (for this example: *The the company tree-bank format is [...]* (replacement NP underlined)).

Another important note also concerns NEs: often NEs are composed of other NEs, as the “Milan” (city name) part in “Inter Milan” (football club name). It is assumed, that NE parts of larger NEs are not referential but rather just label nouns. Aside from the fact that these label nouns are often no NPs and therefore no relevant markables, you must not analyze any NEs any further, even if parts of them should be NPs. Example: do not annotate *Pittsburgh* as a separate markable in *University of Pittsburgh* as *Pittsburgh* is not considered to be referential. Because we annotate [citations as NEs](#), too, you must not divide them into further markables (like author names) either.

Review and Reuse Pre-Annotations

For your convenience we provide you with some pre-annotations in the documents you annotate. These pre-annotations have been made automatically and thus are not always completely reliable. Therefore you should not only add new annotations but also review the existing ones for plausibility and correctness.

Most of the annotations are just pre-annotated markable candidates of which you can simply leave wrong

ones untouched. We will automatically remove all markables later which do not belong to any coreference sets. However, we also provide pre-annotated coreference sets which might be partly or even completely wrong. These wrong coreference sets must not be ignored but they must be corrected by removing elements which are not coreferential with the others. Potentially you will also have to add removed elements to other coreference sets.

When you find and annotate new coreferential markables, you should add them to existing coreference sets where appropriate. Just treat the pre-annotated coreference sets like the ones you have created yourself.

Time, Space and Modality in References

Time, space and modality (cf. [Wikipedia](#)) are constraints which make natural language processing difficult and therefore they also affect your annotations.

Ignoring these constraints is not feasible. Have a look at the following small text to see what would happen if we ignored temporal constraints, for example:

- *In 1969 Peter was the president of ACME. Today his brother is the president*

Peter would be coreferential with *the president of ACME*. The latter would be coreferential with *the president*. The latter would be coreferential with *his brother*. In sum, *Peter* would be coreferential with *his brother*. This is clearly not wanted.

Here are further examples in which time, space and modality constraints apply:

- *In Andorra he is the only Vietnamese.* (spatial constraint *In Andorra*)
- *In 1969 he was the president of ACME Corp..* (temporal constraint *In 1969*)
- *He believes he was the king of Britain.* (modal constraint through *believes*: the person just *believes*, he was the king, which needn't be true – and probably isn't)
- *Our approach should be the best one possible.* (modal constraint through *should*: the authors are unsure whether there is any better approach, otherwise they would have rather used *is* instead of *should be*, i.e., the author's approach is not necessarily *the best one possible*)

In all these examples and for similar cases we would recommend you to not annotate such markables as coreferential.

Unfortunately we cannot give you crystal clear guidance on how to handle such constraints. You should just try to not end up with coreference sets as the one in the introductory example. In this concrete example it would probably make sense to not set *the president of ACME* coreferential with *the president* as these two markables refer to different entities: ACME's president post in 1969 and ACME's president post today. Thus, you should always closely look for spatial/temporal/modal constraints in the context of potentially coreferential markables. Differently constrained markables will often not be coreferential. The maximum context to consider is the whole document but often the context will only be some smaller part of the document like a sentence or even just a phrase.

Tips on Deciding between NP and NE Markable Types

Here are some heuristics with which you can find out whether some markable X is a named entity (NE) or rather just a definite/indefinite noun phrase (NP). None of these is probably sufficient on its own, i.e., you should usually check each of these heuristics before you decide for NE or for NP:

- Try to find out whether there is only one X or if there can be different kinds of X. In the first case this indicates a NE, in the latter case you rather have an NP.
- If X is the name of a conference (e.g., "Coling"), a scientific field (e.g., "Machine Translation"), a system (e.g., "the Stanford Parser"), a technology (e.g., "Dependency Parsing"), a formalism (e.g., "Head-driven Phrase Structure Grammar"), etc., then this indicates a NE.
- If X is capitalized (outside of text parts which are generally capitalized, like section headings), then this indicates a NE.
- If X is in plural or if you can create the plural of X (e.g., "twelve Xs"), then this indicates a normal NP.
- If "another X" sounds strange, then this indicates a NE.
- If X is part of an existential construction (as in "there is X"), then this indicates a normal NP.
- If X is indefinite, then this indicates an NP.
- If X doesn't have a determiner and cannot be used with one (as in "the X"), then this indicates a NE.
- If X is an abbreviation which isn't introduced anywhere in the paper, then this indicates a NE.

If you just cannot decide, then simply annotate the markable as a normal NP.

Some further tips on this topic:

- Abbreviations are often NEs – but *not* always! Especially in the papers that you are annotating, abbreviations are temporarily introduced to save space and typing work. Never just mark an abbreviation as a markable and set it coreferential with all other markables of the same abbreviation just because the abbreviation is the same. Always take the context into consideration and try to find out whether the abbreviation is really a NE or whether it is just a normal NP. In the latter case the abbreviation has to be treated with care, just like any other NP.
- If you are unsure whether some technical term is a named entity (NE) or not, then you should usually treat it just like any other (normal) NP. Most technical terms will be normal NPs. Usually, technical terms that are also NEs are written with a capital letter, i.e., technical terms without a capital letter will usually be normal NPs. There may be exceptions, though, especially for abbreviations (see previous tip). This rule of thumb holds for any NE, by the way. Also, NEs usually have no plural form – although there may be pretty rare NEs which *only* have a plural form, such as “Olympic Games”.
- As stated above, we annotate **only full NP markables**. So, if a NE is preceded by a determiner, then add this determiner to the NE markable as in the first example below. Note, however, that some determiners may create instances of NEs which make the markable NPs (as in example 2). Modifiers may also change NEs to NPs as in example 3. Nonetheless, NEs may also remain NEs despite modifiers: as in example 4 you should always try to find out whether the markable is still a name referring an entity (→ NE) or if it is a description of some entity (→ NP). Examples (relevant markables underlined):
 1. The Eiffel Tower is more than 100 years old. (NE with a determiner)
 2. This is typical for a PCFG. (NP due to an instantiating indefinite determiner)
 3. The Eiffel Tower in Legoland is a smaller replica of the original. (NP due to modifier phrase)
 4. The church in my home town is large. The bigger Eiffel Tower still beats it, though. (NE despite modifier)

Miscellaneous Tips and Clarifications

Here are some random tips and clarifications that should help you with the annotations:

- Indefinite noun phrases (NPs) will usually only be coreferential with markables following the indefinite NPs. So it should not be necessary to look out for indefinite NPs unless you find a markable that refers backwards. So, if you have found a back-referring markable, then you will only need to look for coreferential indefinite NPs *before* this markable, of course. As always, there may be rare counter-examples, which are usually text-type-dependent (relevant coreferential indefinite NPs underlined): *A German soldier was injured by a Greek pacifist. According to media reports, a 20-year-old man threw a stone at the soldier.* Another counter-example which is quite frequent in our text-type (scientific papers) is that conclusion sections (or even other sections of the paper) often reintroduce entities with an indefinite NP although they have already been introduced in the introduction of the paper. In such cases, indefinite NPs may also be coreferential with preceding NPs.
- Appositions where the two elements are coreferential markables can only contain indefinite NPs at the first position if the second element is definite. In other words, the second element of an apposition which contains coreferential markables always has to be definite. Examples:
 - Barack Obama, the president of the United States, announced that ...
 - The president of the United States, Barack Obama, announced that ...
 - Yesterday a dog, my sister's dachshund, lunged at me.
- Actually it should be obvious but nonetheless we'd like to explicitly state this here: NPs with a different number (singular/plural) are usually *not* coreferential. There may be cases where they really are; in your annotations you will probably not find many of these cases, though. So don't hesitate in an example like the following to *not* mark “Wikipedia” and “Wikipedias” as coreferential: *The five largest language editions of Wikipedia are English, German, French, Polish, and Japanese Wikipedias.* Examples of coreferential markables (underlined) with different number:
 - Germany defeated England with 3:0. So the Germans enter the final. (both *Germany* and *the Germans* refer to the German team)
 - The NATO has led many operations and they will probably lead others in the future.
- Reflexive pronouns are just a special case of reciprocal pronouns (see also [Wikipedia entry](#)). Reciprocal pronouns are in almost any case coreferential markables. Besides the **reflexive pronouns mentioned above** there are the following reciprocal pronouns in English: “one another” and “each other”. Reciprocal pronouns should be annotated as definite NPs.
- Elliptic constructions (see also [Wikipedia entry](#)) should be annotated as though they were not elliptic, as long as there is still some part of the markable with omissions left. Examples (valid markables with ellipsis underlined):

- *The green ball is larger than the yellow, while the latter is rounder.*
 - *The red and green ball are lying on the table. The green one is larger than the red.*
 - *Take the rope and pull \emptyset slowly. (\emptyset marks the completely omitted markable; nothing must be annotated here as no part of the markable is left)*
- So-called *identity-of-sense anaphora* (Mitkov 2003, S. 267) bear no coreference. Example: *Some teachers beat their pupils, others hug them*. There is a reference between the two underlined NPs, however, this reference is no coreference as the referred entities are neither the same nor uniquely identifiable.
 - As noted [above](#), variable names are NEs. It should be noted, however, that we only care about variable names, that are used as NPs, i.e., variable names which are part of natural language text. Thus, variable names which are parts of mathematical terms or equations are uninteresting for our annotation task as they are no NPs. A somewhat special case that deserves closer examination is discussed in the following example: in the sentence *a gap's location is specified precisely when $c = d$* the variable names *c* and *d* are considered to be NPs as the mathematical operator $=$ can be seen as an abbreviation for the verb *equals*. You should only take such operators as abbreviations for natural language text, when they are well-known (like $=$, $>$, $+$, etc.) and when they abbreviate simple constructions. If in doubt, then don't annotate variable names used in conjunction with such cases.
 - "Both" can have different meanings depending on its usage and it needs to be differently annotated in different cases of usage. Firstly, as noted [above](#), it can be part of a conjunction. Secondly, it may be a determiner or predeterminer which is just part of some NP (example: *Both men went away.* or *Both her cats are ugly.*); no separate annotation of "both" is required in such cases. Thirdly, "both" may be a pronoun (example: *I want both of them.*) in which case it has to be annotated as a definite NP.
 - In some citations there may be page or section references, e.g., "Smith (2003, p. 21)" or "Smith (2003, section 2.1)". You should just ignore these references and take them as part of the citation if they cannot be left out due to balancing of parentheses. Such references are never full NPs and as such they are uninteresting for our annotations.
 - There may be NPs and possessive determiners which are part of idiomatic expression as in the examples below. Such occurrences are usually not referential and as such no markables. Examples (relevant non-markables underlined):
 - *My God, how could that happen to me? (My is part of the idiom My God)*
 - *John kicked a bucket at Jimmy's head which made Jimmy kick the bucket. (the bucket is part of the idiom kick the bucket)*
 - NPs like "this section" (definite NP), "this paper" (definite NP), "Section 2" (NE), "Figure 12" (NE), etc. are markables, too. While they may be coreferential with each other, they are usually not coreferential with section headings, figure/table captions or any parts thereof. Section headings and captions are not referential and parts thereof are usually no NPs.
 - When it comes to annotating a restrictive relative clause as part of some markable, you should only do so, if the relative clause immediately follows the first part of the markable. Thus, in examples like the following you should not annotate the relative clause as part of the markable (both markable and relative clause underlined):
 - *A car has been developed which is a milestone in technology.*
 - *I saw the car immediately which was buried under a pile of rocks.*
 - Although the word *as* is a preposition, in the term *as such* the word *such* is most probably not an NP and thus no markable for our annotations: the term as such is usually rather an adverb. There will be very few cases – if any – in which the word *such* will be a markable for our annotations.

How to Annotate with MMAX2

Coreference annotation is done using the *MMAX2 annotation tool*. It allows to define markables, to annotate them with a set of specified attributes and to add them to coreference sets where each markable is coreferential with each other.

Installation

To install MMAX2 just extract the distribution archive into a folder of your choice. MMAX2 is written in Java and requires a Java 5 Runtime Environment (JRE). MMAX2 is then started by executing the script `startmmax.bat` (Windows) or `startmmax.sh` (Unix).

Opening MMAX2 Projects

For each paper to annotate there is a corresponding MMAX2 project consisting of several files and

subfolders. The main project file to load from the MMAX2 annotation tool via the *File* → *Load* menu is the one with the .mmax suffix. After loading, MMAX2 prompts the user if he wants to validate the annotation. This is only necessary if the annotation has been manipulated outside of MMAX2, so you can usually choose *Do not validate*.

MMAX2 Windows

MMAX2 consists of three windows: the markable level control panel, the attribute window and the main window, all of which are described in the following sections.

The Markable Level Control Panel

The markable level control panel which is usually opened on the upper right lists all annotation levels. In our annotation task, we have two annotation levels: *coref* for your coreference annotations and *sentence*; the latter is created automatically and is only used for formatting the document (inserting line breaks after each sentence). It is therefore recommended to change the status of the *sentence* level from *active* to *visible* by selecting the corresponding entry in the drop-down menu. This has also the effect that you don't have to explicitly choose the annotation level when working with markables in [the main window](#) since only the *coref* level is active now. After this modification, the markable level control panel is not needed anymore and can be removed by minimizing it or by unchecking the *Show ML Panel* checkbox in the tool bar of [the main window](#).

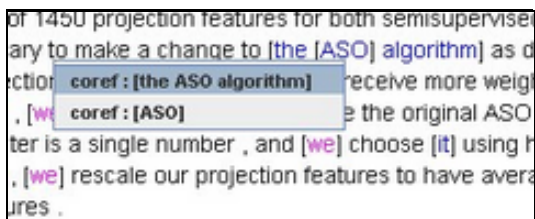


The Main Window

The main window contains the document to annotate. The document already contains preannotated markables that appear in blue or magenta between square brackets. The brackets are called markable handles. Apart from simply visualizing the extent of the markable they are associated with, they are also sensitive to mouse moves: moving the mouse on a markable handle causes the matching handle to be highlighted. Markable handles are also sensitive to mouse clicks and allow direct selection of a markable even in cases of cascaded markables. Once a markable has been selected, it is highlighted with a yellow background.

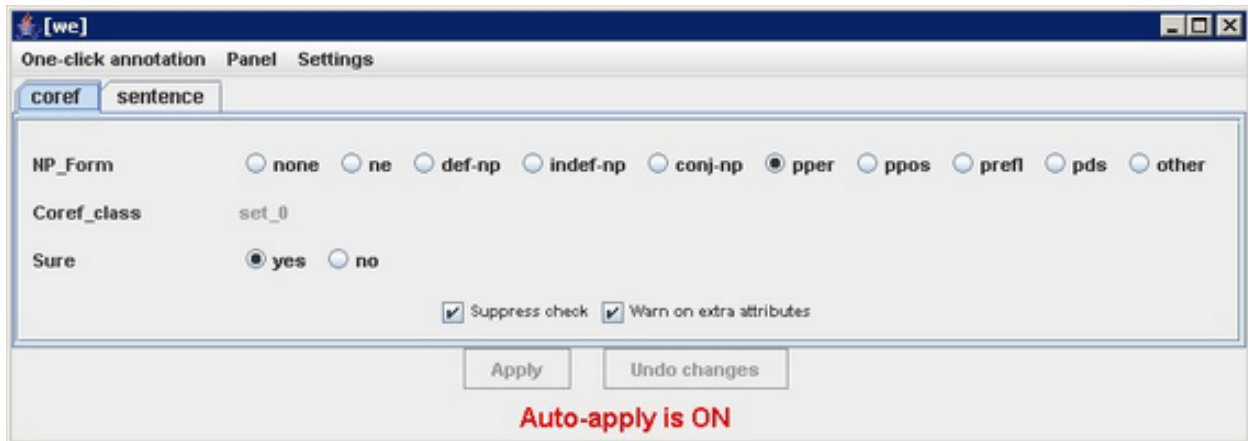
To select a markable for the purpose of displaying its attributes, use the left mouse button. If there is only one markable at a left-clicked position, this markable will be immediately selected and its attributes will be displayed in [the attribute window](#). The same is true if a markable handle is present at the left-clicked position: since each markable handle is associated with exactly one markable, this markable can be immediately selected.

If more than one markable (from one or more levels) is present at a left-clicked position, the click is ambiguous and a pop-up menu containing all markables at this position will be displayed. From this menu select the desired markable by left-clicking its respective entry. The right mouse button, on the other hand, is used for selecting markables for other purposes than display of their attributes, such as [deletion](#), etc. Markable handles and disambiguation pop-up menus, however, work the same with right clicks as for left clicks.



The Attribute Window

Markables are not just sequences of words that have been grouped together; rather, they are the carriers of the actual annotation. Annotations consist of markable attributes and relations. The attributes associated with the currently selected markable are displayed in the MMAX2 attribute window.



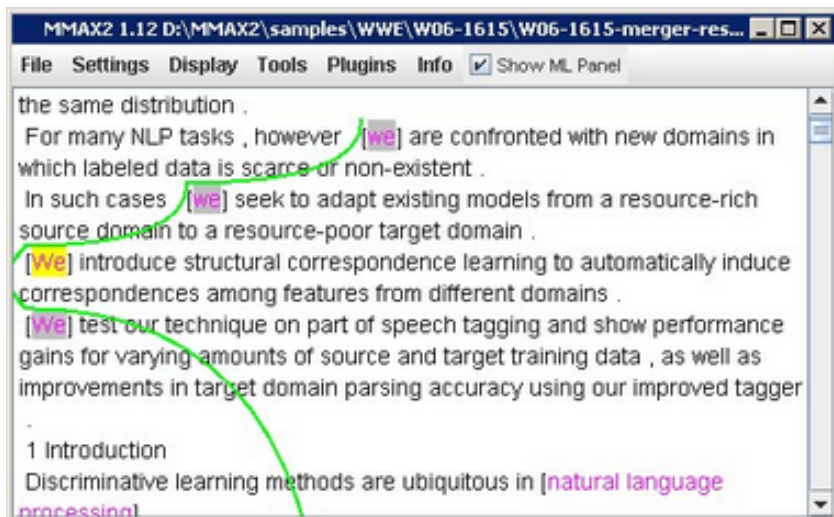
The attribute window contains one tab panel for each markable level in the annotation project. In our annotation project, all annotations are done on the *coref* level, so the *sentence* tab can be ignored.

Attributes consist of a name and a set of possible values, one of which is always selected. Possible values are displayed as radio buttons. Changing the value of a given attribute is done by simply clicking the corresponding radio button. This causes the attribute window to display the newly selected value. By default, the *auto-apply* option is activated, so all changes in the attribute window will immediately (and irrevocably) be applied to the current markable, possibly overwriting earlier values. When using auto-apply, the *Apply* and *Undo changes* buttons are disabled. The *auto-apply* option can be changed in the settings menu in certain situations; usually this will not be necessary, though.

The current annotation scheme contains an attribute *NP_Form* to annotate the type of a markable. It can have one of the following values:

- none: this is the default value. It should always be changed to the correct noun phrase (NP) form. Markables that have this NP form are shown in bold face in the [window](#) to indicate that there is only the default value selected currently.
- ne: choose this for markables that are [named entities](#).
- def-np: choose this for markables that are [definite NPs](#).
- indef-np: choose this for markables that are [indefinite NPs](#).
- conj-np: choose this for markables that are [conjunctions](#) of several NPs.
- pper: choose this for markables that are [personal pronouns](#).
- ppos: choose this for markables that are [possessive pronouns](#).
- prefl: choose this for markables that are [reflexive pronouns](#).
- pds: choose this for markables that are [demonstrative pronouns](#).
- prel: choose this for markables that are [relative pronouns](#).
- other: choose this for markables that are none of the above. This should rarely be the case – if at all.

Apart from carrying annotations in the form of attribute-value pairs, a markable can also be associated with (one or more) other markables to form markable relations. In our annotation task, the *Coref_class* relation is used to represent coreference sets. It can associate any number of markables with each other in a transitive, undirected relation. While markable relations are also attributes in the sense that their availability is controlled by the attribute window, they are visualized graphically in the main window. Selecting a markable that participates in one (or more) relations causes these relation(s) to be graphically rendered by means of colored lines. Additionally, a markable that is a member of a coreference set is rendered in magenta.



Finally, the attribute *Sure* can be used to mark markables where you are not sure if the annotation (either NP form or coreference set membership) is correct. If the attribute is set to *no*, the markable is rendered in red so that it can be found easily in the main window.

Basic Markable Operations

Creating a Markable

If you want to create a new markable, left-click somewhere in the first word, hold the mouse button down and drag the mouse until it is somewhere in the last word. Note that this will only work if no other markable is currently selected! If some other markable should be currently selected, first unselect it by left-clicking on some non-markable or empty space in the display. Upon releasing the mouse button, the selection will automatically be expanded from the beginning of the first to the end of the last word. Also, a pop-up menu will appear with one *Create Markable* menu item for each currently active markable level. Select *Create Markable on level 'coref'* by left-clicking it and a new markable will be created.

The new markable will immediately be rendered in blue and bold. The latter indicates that a non-default NP form still has to be selected. Markable handles, however, will not appear automatically! The display of the contents of the main window has to be refreshed manually to show the handles; this can be done using *Display → Reapply current style sheet* or pressing the shortcut *F5*. Therewith the entire display will be rebuilt and the markable handles become visible on the new markable. Note that markable handles are not required to select a markable, but that they simply offer a convenient alternative. Thus, reapplying the style sheet for every single new markable is not necessary. Instead, it might be good practice (and more convenient) to rebuild the display only once after several new markables have been created.

After creating a new markable, the attribute window immediately shows its attributes that are set to their default values.

Deleting a Markable

If you want to delete an existing markable, select it with the right mouse button. This will cause a pop-up menu to appear. From this menu select the *Delete this Markable* menu item by left-clicking it. Any markable-related rendering styles will immediately be removed from the markable, causing it to be displayed in the default undecorated black font. Also, the handles of the deleted markable will be disabled (indicated with a light gray, strike-through font). With the next refresh of the display (cf. [above](#)) they will be completely removed.

Note that deleting a markable will only work if no other markable is currently selected! In particular this means that you cannot delete the currently selected markable itself. If some markable should be currently selected, then first unselect it by left-clicking on some non-markable or empty space in the display.

It is important to understand that deleting a markable can have an effect on other markables as well: if a markable is the second but last in a coreference set, then deleting this markable will also remove the coreference set.

Resizing a Markable

Resizing a markable means changing the span of words that the markable covers. Words can be added to or removed anywhere in the markable. This means that discontinuous markables are actually supported by MMAX2, however, for our annotation project **they must never be used**.

In order to resize a markable, it has to be the currently selected one.

To add words to the currently selected markable, left-click somewhere in the first word you want to add, hold the mouse button down and drag the mouse until it is somewhere in the last word. Upon releasing the mouse button, the selection will automatically be expanded from the beginning of the first to the end of the last word. Also, a pop-up menu with an *Add to this Markable* menu item will appear. Select this menu item by left-clicking it, and the selected words will immediately be added to the markable. As a result, the added word(s) will be rendered according to the rendering style defined for markables on this level. The markable handle, however, will not be moved to reflect the new markable size until the next display refresh (see [above](#)).

To remove words from the currently selected markable, left-click somewhere in the first word you want to remove, hold the mouse button down and drag the mouse until it is somewhere in the last word. Upon releasing the mouse button, the selection will automatically be expanded from the beginning of the first to the end of the last word. Also, a pop-up menu will appear with a *Remove from this Markable* menu item. Select this menu item by left-clicking it and the selected word(s) will immediately be removed from the markable. As a result, any rendering styles related to the currently selected markable will disappear from the words that have been removed. The markable handle, however, will not be moved to reflect the new markable size until the next display refresh (see [above](#)).

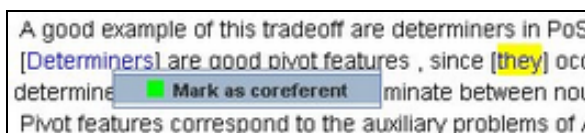
Note that removing all words from a markable is not possible!

Modifying Markable Relations

While markable attributes are modified via [the attribute window](#), markable relations, such as membership in a coreference set, are added and removed by means of mouse actions in the main display, using the right mouse button. Markable relations are always seen with respect to the currently selected markable: if no markable is currently selected, clicking the right mouse button on a markable will not give access to any relation-specific menu items.

Adding Markables to Coreference Sets

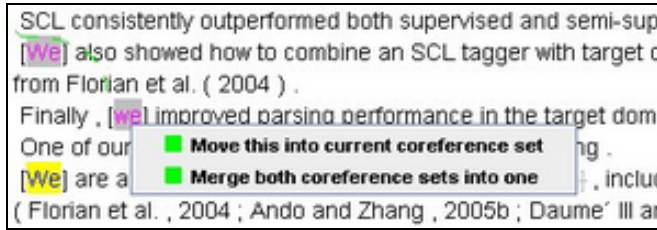
If you would like to put two markables into the same coreference set, then first select one of them as usual and then right-click on the other. If the second markable is not yet participating in a coreference set, a pop-up menu with the *Mark as coreferent* menu item will appear. In addition to the text the menu item will also contain a small box in the same color as used for graphically rendering coreference relations. Select this menu item by left-clicking it, and the second markable will be added to the coreference set of the first markable. As a result, the newly added markable will be graphically linked to its most recent document-order predecessor and successor (unless it is the first or the last markable in its set, respectively). Also, the newly added markable will be rendered in magenta according to its new status with respect to the membership in a coreference set. Finally, the attribute window is also affected by the operation: since adding a markable to a coreference set basically means setting the markable's corresponding *Coref_class* attribute value to the ID of the coreference set, the value for this attribute changes.



If the first selected markable was not previously in a coreference set and it was only through the addition of a markable that a set came into existence, the currently selected markable's rendering style and *Coref_class* attribute will also be changed.

The procedure described above is the most simple case of adding a markable to a coreference set. Things are a little more complex if the markable to be added is already participating in another coreference set. If this is the case, you can decide to adopt the selected markable, or to merge both sets. The set of which the

selected markable is a part is indicated with a dashed green line. In addition, a pop-up menu with two menu items will appear:

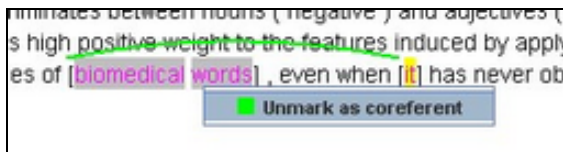


The first menu item will be *Move this into current coreference set*. Selecting this item will adopt the selected markable, i.e., it will be removed from its current and added to the new coreference set. If the coreference set that the “adoptee” markable was part of ceases to exist as a result of the markable’s removal, all set-related attributes and rendering styles will be removed from the left-over markable as well.

The second menu item will be *Merge both coreference sets into one*. Selecting this item will merge the sets of the markable to be added and the set the currently selected markable participates in (if any). Please note that if the markable you first selected is not member in a coreference set, you still get both the *Merge* and *Move* pop-up menu items when you right-click on another markable that is already a member of a coreference set. In case you want to add the first selected markable to the coreference set of the second markable, select the *Merge* item.

Removing Markables from Coreference Sets

You can remove any markable except the currently selected one from a coreference set. If you want to remove a markable from a coreference set, first select another markable of the coreference set and then select the one you want to remove with the right mouse button. A pop-up menu with the menu item *Unmark as coreferent* will appear. In addition to the text, the menu item will also contain a small box in the same color as used for graphically rendering relations of the respective type. Select this menu item by left-clicking it and the markable will be removed from the coreference set. Its set-related attributes and rendering styles will be removed, and the graphical links to other members of the left set will disappear. If the removed markable was the second but last in the set, the set ceases to exist, and set-related attributes and rendering styles will be removed from the left-over currently selected markable as well.



Saving the Annotation

If the annotation has been changed, i.e., after markables or their attributes or relations have been modified, they can be saved by means of the *Save* menu item in the *File* menu. Via the *Levels* submenu, each level can be saved separately. The *All* menu item can be used if more than one level has been modified. Only those markable levels will be saved that actually were modified. A backup copy of the previous version will always be created, but no more than one backup copy will be maintained! You can also set a time limit after which the current annotation is saved automatically. This is done via the *File* → *Auto-Save* menu item.

-
- Mitkov, Ruslan (ed.) (2003): *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.